

Teme licență

Program de burse PPC Renewables

I. Machine Learning și AI

1. Prognoza consumului de energie

a. Definirea problemei și a nivelului de prognoză

Primul pas al proiectului constă în definirea clară a problemei de prognoză. Este necesar să se stabilească, în mod explicit, ce anume urmează a fi prezis - consumul de energie la nivel orar, zilnic sau săptămânal - precum și aria geografică de referință, care poate fi un oraș, o regiune sau o țară. De asemenea, trebuie definit orizontul de predicție, respectiv următoarele 24 de ore, 7 zile sau 30 de zile, precum și granularitatea temporală a modelului, de regulă de o oră. Rezultatul prognozei ar trebui să includă nu doar o valoare punctuală, ci, în mod ideal, și un interval de predicție, de exemplu între percentilele P10 și P90, pentru a reflecta gradul de incertitudine asociat estimării.

b. Colectarea datelor

Etapa următoare presupune colectarea datelor necesare pentru construirea modelului. Este necesar, cel puțin, un istoric al consumului de energie, care să permită o încărcare inițială (initial load) a seriei de timp. În mod opțional, dar recomandat, se pot integra date meteorologice, precum temperatura, umiditatea sau viteza vântului și date privind prețul energiei, în cazul în care se dorește adăugarea unor variabile suplimentare. Ca sursă publică pentru aceste date poate fi utilizată platforma [ENTSO-E Transparency Platform](#).

c. Curățarea, alinierea și verificarea calității datelor

Odată colectate, datele trebuie curățate și aliniate corespunzător. Acest proces implică identificarea și corectarea problemelor legate de fusul orar, tratarea valorilor lipsă și a duplicatelor, precum și investigarea atentă a valorilor extreme (outliers).

d. Explorarea datelor (EDA) și înțelegerea sezonality-ilor

Se recomandă realizarea unei analize exploratorii amănunțite a datelor, cu scopul de a identifica pattern-uri sezoniere - zilnice, săptămânale sau anuale - precum și eventuale anomalii structurale în evoluția consumului.

e. Construirea de variabile predictive (feature engineering)

Pe baza analizei exploratorii, se pot construi variabile derivate, care să capteze mai bine dinamica temporală a seriei, dependențele sezoniere și relațiile cu variabilele externe, precum condițiile meteorologice.

f. Împărțirea corectă a datelor pentru serii de timp

Spre deosebire de problemele clasice de învățare automată, seturile de date de tip serie temporală necesită o metodologie specifică de împărțire între seturile de antrenare, validare și testare, care să respecte ordinea cronologică a observațiilor.

g. Construirea unor modele de referință (baseline)

Este obligatorie construirea unor modele simple de referință, întrucât acestea permit demonstrarea, prin comparație directă, a valorii aduse de modelele mai complexe dezvoltate ulterior.

h. Compararea modelelor candidate

Se recomandă testarea și compararea mai multor modele candidate.

i. Evaluarea performanței cu metrici adecvate

Performanța modelelor trebuie evaluată utilizând metrici specifice seriilor temporale, precum eroarea medie absolută (MAE), eroarea medie pătratică (RMSE) sau eroarea medie procentuală absolută (MAPE), alese în funcție de particularitățile problemei studiate.

j. Interpretabilitatea modelului și analiza erorilor

Este important să se analizeze care variabile au cea mai mare influență asupra predicțiilor, prin intermediul indicatorilor de importanță a variabilelor (feature importance), aplicabili în special modelelor de tip machine learning. De asemenea, se recomandă realizarea unor grafice comparative între valorile prezise și cele reale, alături de o analiză a reziduurilor și a distribuției erorilor. O atenție deosebită trebuie acordată situațiilor în care modelul greșește cel mai frecvent.

k. Livrabilul aplicat

În final, rezultatele proiectului trebuie prezentate sub forma unor grafice clare și ușor de interpretat, care să evidențieze atât performanța modelului dezvoltat, cât și modul în care acesta poate fi utilizat în practică.

2. Prognoza producției de energie

a. Definirea problemei și a nivelului de prognoză

Structura acestei teme urmează aceeași logică descrisă anterior, cu deosebirea că variabila prezisă este, în acest caz, producția de energie, la nivel orar, zilnic sau săptămânal. Este necesar, de asemenea, să se definească zona geografică de interes - oraș, regiune sau țară - orizontul de predicție (24 de ore, 7 zile sau 30 de zile), granularitatea temporală, precum și tipul de output dorit, respectiv o valoare punctuală și, ideal, un interval de predicție între percentilele P10 și P90.

b. Colectarea datelor

Este necesar, în acest caz, un istoric al producției de energie, care să susțină încărcarea inițială a seriei de timp. În mod opțional, se pot adăuga și alte tipuri de date,

precum meteorologice. Sursa publică recomandată pentru aceste date este platforma [ENTSO-E Transparency Platform](#).

c–k. Etapele următoare

Restul etapelor - curățarea și alinierea datelor, explorarea și înțelegerea sezonality-ilor, construirea de variabile predictive, împărțirea corectă a seriei de timp, construirea modelelor de referință, compararea modelelor candidate, evaluarea performanței, interpretabilitatea modelului și analiza erorilor, precum și livrabilul aplicat final - se desfășoară după aceeași metodologie descrisă în cadrul temei anterioare, adaptată la specificul datelor de producție.

II. Data Engineering și Business Intelligence

1. GreenGrid Insights: de la date brute la indicatori de performanță în energia verde

Scopul acestui proiect este construirea unui pipeline de date care colectează, curăță, standardizează și modelează date privind producția de energie regenerabilă, în special energia solară și eoliană. Rezultatul final este un dashboard interactiv, care monitorizează indicatorii de performanță relevanți și oferă informații privind evoluția, variabilitatea și contribuția surselor regenerabile în mixul energetic.

a. Integrarea și automatizarea datelor

Această componentă presupune ingestia datelor din surse publice, precum ENTSO-E sau Open Power System Data (OPSD), urmată de un proces complet de tip ETL (*extract - transform - load*). Procesul ETL trebuie să includă curățarea datelor lipsă și a anomaliilor, alinierea temporală a seriilor de date la intervale orare, precum și normalizarea acestora. Este necesară, de asemenea, implementarea unui mecanism de reîmprospătare automată a datelor, la interval zilnic sau orar, susținut de un sistem de logging și de verificări periodice ale calității datelor, capabil să identifice valorile lipsă sau anomaliile.

b. Modelarea datelor pentru analiză

Pentru a susține analiza ulterioară, este necesară proiectarea unei scheme de date multidimensionale, împreună cu o definire clară și consistentă a indicatorilor de performanță (KPI), documentată printr-un dicționar de date (data dictionary).

c. Vizualizare și suport decizional

Rezultatul final al proiectului constă într-un dashboard interactiv, cu posibilitatea de filtrare pe multiple dimensiuni, capabil să răspundă la întrebări precum: cât a produs energia solară sau eoliană astăzi, comparativ cu media istorică; care este ponderea surselor regenerabile în mixul energetic; când apar scăderi bruște sau anomalii de producție; și cum diferă producția în funcție de sezon sau de ziua săptămânii.

2. PowerTrade Analytics: platformă de analiză și simulare pentru trading în piața de energie

Scopul acestui proiect este construirea unei platforme software de tip trading & decision support pentru piața de energie electrică, care colectează și procesează date publice de piață - prețuri, volume, consum și producție - le modelează pentru analiză și permite simularea strategiilor de tranzacționare (paper trading), evaluarea riscului și monitorizarea performanței traderilor pe piețe precum Day-Ahead (PZU) și Intraday. Platforma urmărește să transforme datele brute de piață în indicatori relevanți pentru decizie, oferind o perspectivă clară asupra volatilității, oportunităților și riscurilor asociate.

a. Integrarea și automatizarea datelor

Această componentă implică ingestia periodică a datelor de piață, incluzând prețuri, volume, consum și producție. Procesul ETL asociat trebuie să asigure curățarea datelor lipsă și a anomaliilor, alinierea temporală a seriilor la intervale orare și normalizarea acestora. Este necesară, de asemenea, o reîmprospătare automată a datelor, la interval zilnic sau orar, susținută de mecanisme de logging și de verificare periodică a calității datelor. Printre sursele publice recomandate se numără [OPCOM](#), [Open Power System Data](#), [Meteostat](#), [Open-Meteo](#) și [EPEX SPOT](#).

b. Modelarea datelor pentru analiză

Modelul de date trebuie orientat spre activitatea de trading, structurat pe piețe, intervale orare și zone. Pe baza acestei structuri se calculează indicatori precum prețul mediu, minim și maxim, volatilitatea, spread-urile dintre piețe sau zone, precum și profitul sau pierderea generată de fiecare strategie. Platforma trebuie să ofere, de asemenea, suport pentru simularea strategiilor de trading, împreună cu reguli simple de gestionare a riscului, precum limitarea pierderilor, limitarea volumului tranzacționat și alertarea automată în cazul unei volatilități ridicate.

c. Vizualizare și suport decizional

Dashboard-ul destinat traderilor trebuie să permită filtrarea datelor în funcție de piață (Day-Ahead sau Intraday), zonă geografică și interval de timp, oferind răspunsuri la întrebări precum: cum a evoluat prețul de astăzi comparativ cu media istorică; care sunt orele cu cea mai ridicată volatilitate; ce profit sau pierdere ar fi generat o strategie simplă de tip cumpărare/vânzare; cum diferă prețurile între zone geografice sau între piețele Day-Ahead și Intraday; și când apar dezechilibre semnificative între producție și consum.